AD

Technical Report ARWSB-TR-13008

NEW PERSPECTIVE ON CLASSICAL ELECTROMAGNETISM

Paul Cote
Mark Johnson

April 2013

ARMAMENT RESEARCH, DEVELOPMENT AND ENGINEERING CENTER
Weapons & Software Engineering Center
Benét Laboratories

Approved for public release; distribution is unlimited

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
25 APR 2013	**FINAL**	**00-00-2011 to 00-00-2012**

4. TITLE AND SUBTITLE **New Perspectives on Classical Electromagnetism**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Paul Cote; Mark Johnson**	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. ARMY ARDEC,Benet Laboratories, RDAR-WSB,Watervliet,NY,12198**	8. PERFORMING ORGANIZATION REPORT NUMBER **ARWSB-TR-13008**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **U.S. ARMY ARDEC, Benet Laboratories, RDAR-WSB, Watervliet, NY, 12198**	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The fallacies associated with the gauge concept in electromagnetism are illustrated. A clearer and more valid formulation of the basics of classical electromagnetism is provided by recognizing a previously overlooked law of induction as well as the physical reality of the vector potential.

15. SUBJECT TERMS
Maxwell?s equations, electromagnetism, Lorenz condition, Coulomb gauge, fallacy.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**21**	

REPORT DOCUMENTATION PAGE	*Form Approved* **OMB No. 0704-0188**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
25-04-2013	FINAL	

4. TITLE AND SUBTITLE New Perspectives on Classical Electromagnetism	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Paul Cote and Mark Johnson	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army ARDEC Benet Laboratories, RDAR-WSB Watervliet, NY 12189-4000	**8. PERFORMING ORGANIZATION REPORT NUMBER** ARWSB-TR-13008
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army ARDEC Benet Laboratories, RDAR-WSB Watervliet, NY 12189-4000	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited or other distribution statement.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The fallacies associated with the gauge concept in electromagnetism are illustrated. A clearer and more valid formulation of the basics of classical electromagnetism is provided by recognizing a previously overlooked law of induction as well as the physical reality of the vector potential.

15. SUBJECT TERMS
Maxwell's equations, electromagnetism, Lorenz condition, Coulomb gauge, fallacy.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Mark Johnson
a. REPORT U/U	**b. ABSTRACT** U	**c. THIS PAGE** U	U	21	**19b. TELEPONE NUMBER (*Include area code)*** 518-266-4986

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI-Std Z39-18

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at lest the year and be Year 2000 compliant, e.g., 30-06-1998; xx-08-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. 1F665702D1257.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. AFOSR-82-1234.

5d. PROJECT NUMBER. Enter al project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORS AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Abstract

The fallacies associated with the gauge concept in electromagnetism are illustrated. A clearer and more valid formulation of the basics of classical electromagnetism is provided by recognizing a previously overlooked law of induction as well as the physical reality of the vector potential.

Keywords

Maxwell's equations, electromagnetism, Lorenz condition, Coulomb gauge, fallacy.

PACS

03.50.De, 41.20.-q, 41.20.Jb

Contents

Introduction

In a critique of conventional physics education, Tony Rothman [1] concludes that despite the impressive contributions of physics in the modern world, physics fundamentals are often presented in a jury-rigged and intellectually dishonest fashion so that the entire enterprise now resembles Bruegel's Tower of Babel. Part of the problem is the perceived need to produce simple instructional results, but these tend to be obtained at the expense of basic understanding of the underlying physics.

Our analysis of the gauge concept in classical electromagnetism suggests that groupthink is another factor in the general problem described by Rothman. The orthodoxy relating to the vector potential, A, has been repeated almost verbatim from one electromagnetism textbook to another. An example of the uncritical acceptance of the current gauge formulation is that the Lorenz gauge has been attributed to H.A. Lorentz for generations. It is only recently that the author of the concept is recognized as L. Lorenz [2]. This correction is made in the latest edition of Jackson's textbook [3].

At present, the gauge concept is the centerpiece of electromagnetism. Electromagnetism is considered a paradigm for gauge theories with the freedom to choose arbitrary functions for a gauge represented as a convenience in problem solution. We show that, on the contrary, the electromagnetic gauge concept is a source of fallacies and confusion that mask fundamental physics principles and phenomena. The gauge approach has evolved into a number of ad-hoc rules that give the right answer: certain types of problem require certain "convenient" gauge choices. A reflection of the overall confusion is the extraordinary number of papers pertaining to this topic.

The conventional view is that only the curl of the vector potential has meaning through its connection to the magnetic field, so the vector potential itself has little physical significance. Konopinski [4] challenged that view. He demonstrates that the vector potential always had physical meaning as the field momentum. His definition is analogous to the definition of the electric field in terms of the force on a test charge q: the field momentum is given by momentum, qA/c, imparted to a test charge, q, as the source current is applied. This gives A measurability, at least in principle. Konopinski's demonstration of the physical reality of the vector potential is important for two reasons. First, it satisfies the normal requirement that physics deals with quantities and relationships that have physical meaning and are testable. The other reason is that once the physical reality of the vector potential is recognized, the notion that one can always add any arbitrary scalar component is no longer tenable because there are restrictions relating to the fact that the source of this component must now be physical in origin.

Another misconception is that only the vector potential is undefined. Actually, in the current formulation, in the presence of dynamic Coulomb fields, none of the fields is completely defined if the vector potential is left undefined. These issues are resolved in the following by recognizing that the vector potential is determined by both the magnetic field and the induced electric field.

Demonstration of gauge fallacies

A contributing factor to the gauge-related fallacies is the failure to distinguish among variables. In the standard formalism, the labels E, A, and φ are each commonly used to represent a variety of different variables. We use subscripts throughout the following as an aid to clearly preserving the meaning of a given variable.

We begin with some elementary notes. According to the Helmholtz theorem, any physically meaningful vector can be written as a sum of a gradient of a scalar and a curl of a vector. A gauge choice is required if one needs to obtain quantitative expressions for variables that are described only by a curl equation. One generally chooses a gauge that produces the simplest, least cumbersome form for relationships among variables, in analogy with the choice of zero for the Coulomb potential. The choice of zero for the gauge usually serves this purpose. Alternative choices of gauge cannot affect the fundamental physics.

Hidden gauge

The magnetic field B is given by,

$$B = \nabla \times A . \tag{1}$$

According to the Helmholtz theorem, the general expression for the vector potential, A, is given by

$$A = \nabla \times F_A + \nabla \varphi_A , \tag{2}$$

so Eq. (1) defines A to within an arbitrary function, $\nabla \varphi_A$, which means that a gauge choice is required if no other information is available. A non-zero divergence of a vector implies the existence of a scalar field associated with that vector. The following summarizes the usual development of the gauge based approach and illustrates how it is a source of confusion.

Faraday's law is originally expressed as a line integral of the induced field, E_I, around a closed path, which leads to the relationship, $\nabla \times E_I = -\partial B / \partial t$. Since the Coulomb field is always derivable from a scalar potential, $E_C = -\nabla \varphi_C$. (In the general case, φ_C is a retarded potential, as discussed in section D.) When these two electric fields are present together, the total field is $E = E_I + E_C$, giving $\nabla \times E = -\partial B / \partial t$. Applying Eq.(1) gives $\nabla \times (E + \partial A / \partial t) = 0$. Thus, $E + \partial A / \partial t = \nabla \varphi$, with $\nabla \varphi$ representing the gradient of a scalar potential so that E is given by,

$$E = -\nabla \varphi - \partial A / \partial t . \tag{3}$$

This approach presents two problems. First it artificially couples the two basic electric fields as if the principle of superposition no longer applies. Second, it leaves the impression that A is the only variable requiring a gauge choice. Regarding the first item, this artificial coupling of the two very different electric fields is used later as justification for introducing the Lorenz condition to effect a decoupling [3,5,6]. Decoupling is needed to address a wide variety of problems. Regarding the second item, if one preserves the distinction among variables, it is seen that there are actually two gauges to consider. A clearer development of the basic equations is offered next.

Adhering to the original form of the Faraday law, which relates only to the induced field, E_I , one obtains

$$\nabla \times E_{\mathrm{I}} = -\partial B / \partial t = -\nabla \times \partial A / \partial t \, , \tag{4}$$

by applying Eq.(1). As with Eqs. (1) and (2), the curl provides an incomplete definition of E_I since

$$\nabla \times E_{\mathrm{I}} = \nabla \times (E_I + \nabla \varphi_I) \, . \tag{5}$$

So, the general expression for E_I is,

$$E_I = -\partial A / \partial t - \nabla \varphi_I \, . \tag{6}$$

Therefore, the general expression for total electric field is,

$$E = E_C + E_I = -\nabla (\varphi_C + \varphi_I) - \partial A / \partial t \, , \tag{7}$$

with A given in general form by Eq.(2). Comparing Eqs.(3) and (7) shows that, in general, φ is not the Coulomb potential as conventionally assumed, but is the sum of two scalar fields, $\varphi = \varphi_C + \varphi_I$. So the common textbook assumption that φ is just the Coulomb potential represents the adoption of a hidden gauge choice, $\varphi_I = 0$. Similarly, Eq.(6) shows that the standard practice of employing

$$E_I = -\partial A_S / \partial t \tag{8}$$

implies the same hidden gauge choice. The subscript on A_S is used to denote the adoption of the standard gauge. Each gauge choice results in a different, but physically meaningful, vector potential, as we shall see. Note that φ_I =0 does not necessarily mean $\nabla \cdot E_I = 0$ because the vector potential is still described by Eq.(2). *This hidden gauge choice is always made in electromagnetism in either of the two ways given above, so* φ_I *=0 is always the implicit "standard gauge" in electromagnetism.*

Hidden gauge and Gauss' law

Gauss' law for the basic dynamic E_C and E_I fields is given by

$$\nabla \cdot E = \nabla \cdot (E_C + E_I) = \rho / \varepsilon \, . \tag{9}$$

Inserting Eq.(7) into Eq.(9) gives the general expression for Gauss' law,

$$\nabla \cdot E = -\nabla^2 (\varphi_C + \varphi_I) - \partial (\nabla \cdot A) / \partial t = \rho / \varepsilon \, , \tag{10}$$

where A is given by Eq.(2). Applying the hidden gauge, φ_I =0, gives

$$-\nabla^2 \varphi_C - \partial (\nabla \cdot A_S) / \partial t = \rho / \varepsilon \, . \tag{11}$$

Since $\nabla \cdot A_S = \nabla^2 \varphi_A$, Eq.(11) can be rewritten as

$$\nabla^2 (\varphi_C + \partial \varphi_A / \partial t) = \nabla^2 \varphi = -\rho / \varepsilon \, . \tag{12}$$

We now consider the corresponding Coulomb gauge ($\nabla \cdot A_C = 0$, $\varphi_A = 0$) expression to complete the results for the two simplest gauge choices ($\varphi_I = 0, \varphi_A = 0$). From Eq.(10),

$$\nabla^2 (\varphi_C + \varphi_I) = \nabla^2 \varphi = -\rho / \varepsilon \, . \tag{13}$$

Note the similarity of Gauss' law in the two gauges, Eqs.(12) and (13). Both express the same physics contained in the hidden law, but in different forms. Both say that the sum of two dynamic scalar fields obeys Gauss' law. Consequently, the same electrostatic Laplacian expression applies to sum of the dynamic potentials in both gauges. As we will show, this result is actually a basic feature of retarded fields.

There is a simple relationship between these two gauge scalar potentials which is obtained by equating the Gauss' law expression in the two respective gauges, the standard gauge, $\varphi_I = 0$ and the Coulomb gauge, $\varphi_A = 0$. The result is,

$$\varphi_I = \partial \varphi_A / \partial t \, . \tag{14}$$

We will revisit Eq. (14) to show it is an identity, rather than an equality, because both terms represent the same induced scalar potential.

Hidden gauge and the peculiarity of the Coulomb gauge for dynamic fields

Jackson's textbook [3] offers a demonstration that the standard treatment of the gauge concept leads to the requirement that dynamic Coulomb fields propagate instantaneously. He describes this as a peculiarity.

We reproduce Jackson's demonstration here. First, ignore the distinction between A_C and A_S. Next, apply $E_I = -\partial A / \partial t$. Finally, invoke the Coulomb gauge, $\nabla \cdot A = 0$, in the dynamic form of Gauss' law, Eq.(9). This gives,

$$\nabla^2 \varphi_C = -\rho / \varepsilon \, . \tag{15}$$

Since φ_C is a basic field, and cannot be considered a sum of more fundamental fields in the manner described in section B, Eq.(15) indeed requires that dynamic Coulomb fields propagate instantaneously, which is an absurdity. The resolution of this issue is addressed next.

Overlooked law of physics

In the usual derivation of the dynamic form of Gauss' law, Eq.(9), involves nothing more than inserting the sum of the dynamic fields to the static expression for the divergence of the Coulomb field. That is not a legitimate procedure. Closer examination of Eq. (9) reveals an overlooked physical law. Under dynamic conditions the Coulomb field can no longer treated as propagating instantaneously, so it no longer obeys Gauss' law. The overlooked law can be expressed as: *A dynamic Coulomb field induces a self-correcting scalar component for the induced field so that the total electric field obeys Gauss' law.* Consequently, E, which is the now sum of retarded Coulomb and induced fields, appears to propagate instantaneously from a central source or a distribution of central sources, as required by Eq.(9). In other words, E must behave as if it were propagating as a quasi-static field with all field lines originating and terminating on the instantaneous position of the moving charges.

The basic example of this phenomenon is the pair of Lienard-Weichert retarded potentials for a moving point charge. For a moving charge, neither E_C nor E_I at a distant point in space can originate from the present location of the moving charge because of the finite speed of light. The total E field, although no longer spherically symmetric, is nevertheless radial and tracks the instantaneous position of the charge, so E obeys Gauss' law. Thus, despite the fact that both fields propagate at the speed of light, the sum, E, at a distant point is directed as if it were an instantaneously propagating field. This effect is arguably the most fascinating phenomenon in electromagnetism. The principal reason that its central role in the basic phenomena has been ignored to date is that the present gauge approach masks its presence.

As for the dynamic form of Gauss' law, the need for Eq. (9) and the overlooked law rests on charge conservation. This can be seen by taking the time derivative of Gauss' law, Eq.(9), which gives,

$$\nabla \cdot J_T + \partial\rho / \partial t = 0 . \tag{16}$$

where J_T is the true current and the solenoidal total current, $J_{TOT} = J_T + \varepsilon\, \partial E / \partial t$. So the real basis for the dynamic form of Gauss' law is the requirement of charge conservation.

The overlooked law also provides the mechanism that explains the remarkable fact that all source currents in electromagnetism propagate instantaneously, no matter how rapidly these currents may vary. How does nature accomplish this without violating relativity? The explanation is that the displacement current fields and the fields that drive the charges in a conductor are always a sum of basic retarded E_C and E_I fields, and this sum appears to propagate instantaneously, as described above (if one ignores field distortions).

Returning to the overlooked law, the situation for E_I is similar to that of the vector potential. Faraday's law offers an incomplete description of the induced field because it only defines E_I via a closed line integral. Consequently, any scalar component is left undefined. A complete definition of E_I is available using a precise mathematical expression of the overlooked law that is, in effect, lying in plain sight: recognizing the scalar wave equation for the Coulomb field as a fundamental requirement of Coulomb fields (which we will return to later),

$$\nabla^2 \varphi_C - (\partial^2 \varphi_C / \partial t^2) / c^2 = -\rho / \varepsilon , \tag{17}$$

and comparing this with the general expression for Gauss' law, Eq.(9) shows

$$\nabla \cdot E_I = (\partial^2 \varphi_C / \partial t^2) / c^2 . \tag{18}$$

With Eq.(18) accepted as a fundamental equation, one now has a complete characterization of E_I as well as E_C (via Eq.(9)). The crucial example is the one given above, since Eq.(18) is required by the wave equation for φ_C. The basic physics equations must require that Coulomb fields propagate at the speed of light, regardless of any gauge choice. By contrast, as shown in section C, the present textbook formalism leads to the conclusion that Coulomb fields propagate instantaneously in the Coulomb gauge.

Combining Eq.(18) with the standard gauge expression, $E_I = -\partial A_S / \partial t$, gives

$$\nabla^2 \varphi_A = \nabla \cdot A_S = -\,(\partial \varphi_C / \partial t) / c^2 . \qquad (19)$$

Equation (19) appears identical to the familiar Lorenz condition, but now it is required by a law of physics and is not just a convenience that provides the right answer. In physics, as in medicine, a condition is something to be avoided, so we will refer to Eq.(19) as the Lorenz equation since it is now a valid equation and not an arbitrary function. Equation (19) is actually the field momentum expression of the overlooked law, Eq.(18); both Eqs.(18) and (19) have the dynamic Coulomb field as the physically meaningful common source. (It might be argued that the addition of a static scalar potential to Eq.(19) will also satisfy Eq.(18). This possibility is dismissed as physically meaningless, however. Also, according to Konopinski's definition of A, such static sources are irrelevant.)

The conclusion regarding the role of retarded fields and current continuity in establishing the physical necessity of Eqs.(17), (18), and (19) was developed using the argument that the basic Coulomb fields must propagate at the speed of light. We can arrive at the same conclusion more directly using the pair of general retarded field equations [3,5,6] for the vector and Coulomb potentials.

First, we provide a new perspective on these equations. Equations (20) and (21) are presented in one textbook after another as solutions to the wave equations that are derived from Maxwell's equations with the incorporation of the Lorenz condition, which, as we have discussed, is only valid in the standard gauge. So, in order to show that Coulomb fields, for example, propagate at the speed of light, one needs to invoke an arbitrary function, the Lorenz condition, which is another absurdity. Moreover, this approach is misleading because these equations apply more generally and are not restricted to the standard gauge, as we will demonstrate in Section F. It can be seen by inspection that they are straightforward modifications of the corresponding static expressions using only the requirement that fields must propagate at the finite speed of light. They are therefore first principles results that enable us to avoid the arbitrary Lorenz condition:

$$A(1,t) = \int \frac{\mu\, j(2,t_r)}{4\pi r_{12}} dV_2 , \qquad (20)$$

and,

$$\varphi_C(1,t) = \int \frac{\rho(2,t_r)}{4\pi\varepsilon r_{12}} dV_2 , \qquad (21)$$

where the fields are evaluated at position 1 and present time t, from sources located at position 2 and retarded time $t_r = t - r_{12} / c$.

Using Eqs.(20) and (21) as the starting point, one can now directly obtain the wave equation for $\varphi_C(1,t)$ and the Lorenz equation from first principles by differentiation. This involves using the chain rule and $\nabla_1^2(1/r_{12}) = -4\pi\delta(r_{12})$ for the scalar wave equation, and the chain rule, $\nabla_1 f(r_{12}) = -\nabla_2 f(r_{12})$, and integration by parts, where surface integrals of physically real variables vanish at large distances, for the Lorenz equation. Equation (21) and, thus, the wave equation, are independent of gauge choice because Eq.(21) is independent of the vector potential. It follows that Eq.(18) must apply because of Gauss' law.

We now have that the vector potential, A_S , is completely defined in the standard gauge by Eqs.(2) and (19). Equation (19) also illustrates the point that physics dictates the choice of the second gauge, φ_A, once the first gauge choice is exercised ($\varphi_I = 0$). This clash of gauge choices is the main flaw in Jackson's demonstration in section C. The above completely undermine the whole premise of the gauge approach which holds that one is always free to make arbitrary gauge choices, and it provide a physical explanation for the apparent central role played by the Lorenz equation. We note that the existence of a connection between the Lorenz equation and the continuity condition has already been stated in textbooks (e.g., Feynman[5] , Panofsky and Phillips[6]).

In summary, by explicitly adopting the previously unrecognized standard gauge, $\varphi_I = 0$ (via Eq.(8)) and adding Eq.(18) to complete the list of Maxwell's equations, we now have a complete set of fundamental equations. All variables have physical meaning and measurability and are completely defined. The focus on the electromagnetic gauge concept and the Lorenz condition has been replaced by a simpler and physically more meaningful formulation centered on recognition of the fundamental role of retardation effects.

As for the gauge concept, one can initially fix the induced field gauge (φ_I =0) in a footnote, and forget it. (Something resembling this is done in practice, as reflected in the frequent use of the Lorenz condition, but it is framed in an invalid gauge context.) As will be made evident, once the standard gauge is adopted, the alternative Coulomb gauge approach is primarily useful as an instructional device. We continue illustrating the implications of these results, in the following sections.

In the absence of dynamic Coulomb fields

Consider cases where dynamic Coulomb fields are absent, so that there are no retarded Coulomb fields. The textbook treatments of these cases generally invoke $\nabla\cdot A_C = 0$ ($\varphi_A = 0$) as the convenient gauge, with labels such as Coulomb gauge or radiation gauge.

As discussed earlier, a dynamic Coulomb field is the only possible, physically meaningful source of a scalar component for induced fields or for field momentum. (i.e., Eqs.(18) and (19)). In the absence of any dynamic Coulomb field, induced scalar fields have no meaning, so physics requires

$\nabla \cdot A = \nabla \cdot A_C = \nabla \cdot A_S = 0$. Consequently, in the absence of dynamic fields, the gauge concept itself is meaningless.

Note that even without our introduction of the overlooked law, or Konopinski's definition of the vector potential, the gauge approach cannot apply in this case. Absent Coulomb fields, Gauss' law always requires that the induced field be solenoidal. In view of Eq. (8), there never was a free choice for $\nabla \cdot A$.

As an elementary example, consider a circular metal ring in a uniform, time varying magnetic field (Fig. 1a). This is similar to the betatron configuration. Recall that potential differences are given by a line integral of E_C, while emf's are given by a line integral of E_I (which also illustrates the point we made earlier regarding the need to maintain the distinction among fields for even the most elementary problems). Given the ring symmetry and field uniformity, there can be no potential differences, so the apparent instantaneously transmitted current changes are actually generated by the uniform distribution of an emf along the ring. Thus, there is no violation of relativity, no induced scalar potentials, and, thus, no free choice for $\nabla \cdot A$. The vector potential A must be solenoidal in this case.

The same holds for electromagnetic radiation in free space, remote from any sources. So there are no Coulomb fields, and, again, only solenoidal fields exist. In classical electromagnetism, the radiation fields and their sources propagate through space together at the speed of light by means of successive generations of time varying closed loops of magnetic fields which induce closed loops of displacement currents, which, in turn, induce new magnetic field loops, continuing ad infinitum. Despite this, electromagnetic radiation is always addressed in terms of the convenient choice of the radiation gauge, $\nabla \cdot A = 0$, as if there were a choice.

Wave equations in the two gauge choices

We now consider the wave equations in the two simplest gauge formulations ($\varphi_I = 0$, $\varphi_A = 0$). The general expression for wave equation for the vector potential in terms of the two arbitrary gauge choices is obtained using Ampere's law,

$$\nabla \times B = \nabla \times \nabla \times A = \mu J_{TOT} = \mu(J_T + \varepsilon\, \partial E / \partial t)\,, \tag{22}$$

where J_T is the true current, and $\varepsilon\, \partial E / \partial t$ is the sum of the two displacement currents. Applying Eq.(7) to Eq.(22) gives the general wave equation (prior to any gauge selection):

$$\nabla(\nabla \cdot A) - \nabla^2 (A) = \mu J_T - \mu\varepsilon\, \partial \nabla(\varphi_C + \varphi_I) / \partial t - \mu\varepsilon\, \partial^2 A / \partial t^2\,. \tag{23}$$

The vector potential wave equation in the standard gauge, $\varphi_I = 0$, is obtained directly from Eq.(23), giving the familiar standard form for the wave equation for the vector potential, A_S,

$$\nabla^2 A_S - \mu\varepsilon\, \partial^2 A_S / \partial t^2 = -\mu J_T + \nabla(\nabla \cdot A_S + (\partial \varphi_C / \partial t) / c^2)\,. \tag{24}$$

Thus, whenever these familiar expressions are employed, one must recognize the implicit assumption of the standard gauge choice.

Applying the Lorenz equation, Eq.(19), gives the familiar wave equation for the non-solenoidal A_S,

$$\nabla^2 A_S - \mu\varepsilon\,\partial^2 A_S / \partial t^2 = -\mu J_T \,. \tag{25}$$

A key point is that the general solution to Eq.(25) is given by the retarded field integral in Eq.(20) over the source current, which, in the present case, is the true current , J_T . Typically, true currents refer to moving charges in conductors. In a circuit that contains a Coulomb field, e.g., Fig. 1b, the portion of the circuit represented by J_T is incomplete, and therefore non- solenoidal. The vector potential, A_S, generated by that source is obtained by the integral over J_T using the general solution for A_S in Eq.(20). As indicated in our earlier discussion, if one calculates the divergence of A in Eq.(20), one finds that solenoidal sources ($\nabla\cdot j = 0$) produce solenoidal fields and non-solenoidal sources ($\nabla\cdot j \neq 0$) produce non-solenoidal fields. So the physical meaning of the vector potential obtained in the standard gauge is that A_S is a component of the total vector potential that would be obtained from the entire solenoidal current loop. Hence, it cannot be solenoidal.

We now discuss the corresponding wave equations in the Coulomb gauge formulation. Eq.(8) no longer applies for E_I . Returning to the general expression for E_I , Eq.(6), with $\varphi_A = 0$ gives,

$$E_I = -\partial A_C / \partial t - \nabla\varphi_I \,. \tag{26}$$

In the Coulomb gauge formulation, the corresponding wave equation from Eq.(23) is

$$\nabla^2 A_C - \mu\varepsilon\,\partial^2 A_C / \partial t^2 = -\mu J_T + \mu\varepsilon\,\partial\nabla(\varphi_C + \varphi_I) / \partial t \,. \tag{27}$$

The scalar wave equation for φ_C obtained from Eq.(21) applies here because is independent of gauge choice. It now requires,

$$\nabla^2\varphi_I = \nabla\cdot E_I = -\partial^2\varphi_C / \partial t^2 / c^2 \,. \tag{28}$$

Thus, the overlooked, fundamental law of induction, Eq.(18), is obtained in both formulations which satisfies the criterion of gauge invariance for fundamental equations.

Returning to Eq. (26), it describes how the dynamic Coulomb field now induces the scalar φ_I directly in the expression for E_I in the Coulomb gauge formulation, rather than indirectly via the vector potential, as in the standard gauge. And it explains more clearly why Eq.(14) is actually an identity.

As for the physical meaning of the vector potential in the Coulomb gauge, note that the displacement current that completes the current loop with J_T arises from the time derivative of $\nabla\varphi$ in Eq.(27) , which obeys Gauss' law at all times. Thus, it behaves as if it were an instantaneously propagating longitudinal field, producing an instantaneously propagating displacement current. The complete current loop serves as the solenoidal current source for the solenoidal vector potential, A_C. In this case, the solution for A_C is obtained from Eq.(20) using $J_T - \varepsilon \partial \nabla \varphi / \partial t$ as the source current.

The wave equations in the two gauge contexts, Eqs. (25) and (27), express the same physics despite the fact that the Lorenz equation does not apply in the Coulomb gauge. The divergence of both returns the continuity equation, and the curl of both returns the wave equation for the magnetic field, B. (The reason that the two different vector potentials A_S and A_C give the same B is that there is no net contribution from the scalar component.) Both vector potential wave equations give the wave equation for the induced field, E_I. And, one can complete the set of basic equations with the gradient of Eq. (17), which gives the wave equation for E_C, which is valid for both gauges.

Finally, it should be noted that the freedom to assign the induced scalar field between φ_A and φ_I is directly related to the Lorenz transformation where one is free to add a gradient of a scalar to the vector potential as long as one subtracts a compensating term from the scalar potential. Thus the Lorenz transformation is more than a mathematical trick as usually portrayed, it has physical significance. Furthermore, as illustrated by the two simplest gauge choices above, the physical meaning of the variables is maintained with this transformation if one preserves the identity of the variables.

Four- vector formulation

The ability to express the basic equations in four-vector form is conventionally used to justify the otherwise arbitrary application of the Lorenz condition. In the following, we modify the standard four-vector notation to more precisely reflect gauge choice, and also show that the Lorenz equation and the standard gauge are not unique in providing a four-vector formulation. (The requirement for the four-vector formulation follows from the fact that Maxwell's equations are invariant to a Lorentz transformation.)

Following the $c = 1$ notation in Feynman et al [5], the four- vector has the form,

$$a_\mu = (a_t, a_x, a_y, a_z). \tag{29}$$

The four-vector divergence is given by

$$\nabla_\mu a_\mu = \partial a_t / \partial t + \nabla \cdot a, \tag{30}$$

and, the four-vector Laplacian is given by

$$\Box^2 = \partial^2 / \partial t^2 - \nabla^2. \tag{31}$$

Preserving the distinction among vector potentials, the standard gauge potential is given by,

$$(A_S)_\mu = (\varphi_C, A_S)\,. \tag{32}$$

Using the current vector,

$$j_\mu = (\rho, (J_T)_x, (J_T)_y, (J_T)_z)\,, \tag{33}$$

gives,

$$\Box^2 (A_S)_\mu = j_\mu / \varepsilon\,, \tag{34}$$

and,

$$\nabla_\mu \, j_\mu = 0\,. \tag{35}$$

Equation(19), the Lorenz equation, is given by,

$$\nabla_\mu (A_S)_\mu = 0\,. \tag{36}$$

To illustrate that the Coulomb gauge results can also be expressed in four-vector form, the Coulomb gauge formulation developed earlier gives:

$$\Box^2 (A_C)_\mu = j_\mu / \varepsilon\,, \tag{37}$$

where,

$$(A_C)_\mu = (\varphi_C, A_C)\,, \tag{38}$$

and,

$$j_\mu = (\rho, (J_T - \varepsilon\, \partial \nabla \varphi / \partial t)_x, (J_T - \varepsilon\, \partial \nabla \varphi / \partial t)_y, (J_T - \varepsilon\, \partial \nabla \varphi / \partial t)_z)\,, \tag{39}$$

where φ is the sum of scalar potentials. In this case,

$$\nabla_\mu \, j_\mu = \partial \rho / \partial t\,. \tag{40}$$

Thus, we have shown that the Coulomb gauge formulation produces all the basic equations of electromagnetism, complete with their four-vector expressions, where the Lorenz equation does not apply. So the idea that the Lorenz condition, (i.e., the standard gauge) is essential to electromagnetism can be included on the list of fallacies. The one factor that is fundamental in both the standard and the Coulomb gauge formulations is the previously overlooked law of induction, Eq. (18).

Picture of reality

Many of the points made in Part II are summarized in schematic form in Figure 1. It compares the two general categories of problems encountered in electromagnetism. Figure 1a represents cases where external dynamic Coulomb fields are absent and Figure 1b represents cases where dynamic Coulomb fields are present. In Figure 1a, either the true current, J_T, or the displacement current $J_D = \varepsilon \partial E / \partial t$ forms a closed loop so that the resulting vector potential field (or field momentum) is solenoidal and the gauge concept does not apply ($A = A_S = A_C$). In the case of a current induced in a metal ring by a uniform magnetic field, the emf is uniformly distributed along the ring and the fact that current changes occur simultaneously at all points along the ring does not violate relativity.

A related point is that the symmetry of the ring in a changing magnetic field does not suspend Ohm's law, so that internal potential differences are, in fact, present in any given ring segment from that effect. These potential differences are exactly cancelled by the charge distribution generated by E_I so that no net charges exist anywhere along the ring.

Consider the case where a steady state AC current is generated by a localized emf in the ring. Now the instantaneously transmitted current changes provide an illusion of violating relativity. In this case, however, a uniformly induced emf arises from the self generated magnetic field, so the net emf acts in the same way as in the previous example to preserve charge continuity. Again the E_I fields play an essential role.

Figure 1b corresponds to more typical circuits where both types of current, J_T and an external J_D exist together to form a closed loop ($A_C \neq A_S$). The overall circuit geometry is similar, as shown in the two figures, so the fields are similar in both. It should be clear from this diagram that the Coulomb gauge vector potential A_C is unique, solenoidal, and thus completely defined in both figures. As discussed, the A_S' field is a non-solenoidal component of A_C'. (A prime is added to the field terms to reflect the expected differences in actual current configurations between the circuits in Figures 1a and 1b.)

An important difference between Figures 1a and 1b is that the E_C field around the capacitor in Figure 1b can extend throughout space. This is actually a complicated retarded field generated by the moving charges on the capacitor plates. To understand even these most elementary problems, one needs at least an awareness that retardation effects are at the heart of electromagnetism. The basic feature of retarded fields is that their sums, $E = E_C + E_I$, are always similar to quasi-static fields. These E fields originate and terminate at the instantaneous position of the charges on the capacitor plates, so that the continuity condition applies at all times.

Summary and Conclusions

The present analysis provides the simple modifications needed to remove the gauge fallacies and reveal the basic role of retardation effects. It also shows that the present imprisonment of electromagnetism in Bruegel's tower need not be a life sentence. The major conclusions regarding the gauge fallacies in electromagnetism are listed in the following:

a. The relationship, $\nabla \cdot E_I = (\partial^2 \varphi_C / \partial t^2) / c^2$ is a previously unrecognized fundamental law of induction in electromagnetism. It is a tacit recognition of the central role of retardation effects and supplements Faraday's law of induction, to permit a full definition of E_I , as well as E_C . It completes the set of Maxwell's equations and can properly be termed a missing Maxwell equation. Furthermore, the Lorenz equation is more than a mere convenience. It is an alternative statement of the previously missing law of induction in the gauge, $\varphi_I = 0$, which provides physically meaningful characterization of the field momentum in that gauge; on the other hand, it only meaningful in that gauge.

b. Generally, two gauge choices are required rather than one. The previously hidden choice, $\varphi_I = 0$ must be recognized. Absurdities occur when conflicting gauge choices are adopted.

c. As shown by Konopinski, the vector potential has both physical meaning and measurability as a field momentum.

d. In the absence of dynamic Coulomb fields, the gauge concept is invalid; A is always uniquely defined because $\nabla \cdot A = 0$ is a physical requirement resulting from the absence of scalar field sources.

e. The Coulomb gauge does not require that dynamic Coulomb fields propagate instantaneously. That "peculiar" conclusion originates from the erroneous, simultaneous adoption of conflicting gauge choices.

f. The gauge concept only applies in the special case where dynamic E_C fields exist, and, in that case, it only applies under the restrictions of physical laws. Thus, it is inappropriate to characterize electromagnetism as a gauge theory paradigm.

References

1. T. Rothman, The man behind the curtain, American Scientist, **99**(3), May-June 2011.

2. L. Lorenz, On the identity of vibrations of light with electrical currents, Philos. Mag. **34**, 287, 1867.

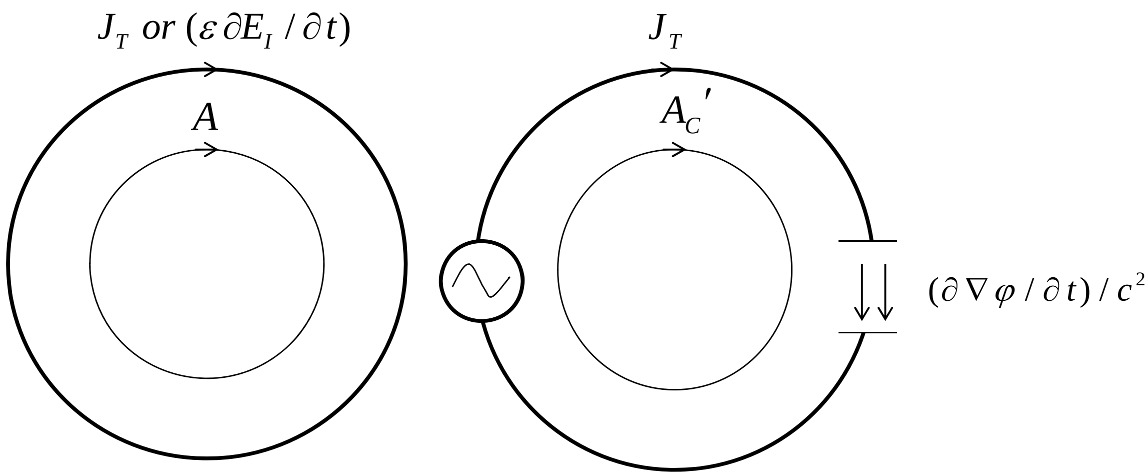
3. J.D. Jackson, Classical Electrodynamics, 3rd Ed. (John Wiley and Sons, New York, 1999).

4. E.J. Konopinski, What the electromagnetic vector potential measures, Am. J. Phys., **46** (5) 499-502, 1978.

5. R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures in Physics vol II* (Addison-Wesley, Reading, MA, 1964).

6. W.K.H. Panofsky and M. Phillips, Classical Electricity and Magnetism, 2nd Ed. (Addison-Wesley , Reading, MA, 1962).

APPENDIX A



$$A = A_C = A_S = \nabla \times F_A$$

$$E_I = -\partial A_C / \partial t = -\partial A_S / \partial t$$

$$B = \nabla \times \nabla \times F_A$$

Figure 1a

$$A_C' = \nabla \times F_A'$$

$$A_S' = \nabla \times F_A' + \nabla \varphi_A$$

$$E_I' = -\partial A_C' / \partial t - \nabla \varphi_I = -\partial A_S' / \partial t$$

$$B' = \nabla \times \nabla \times F_A'$$

Figure 1b

Figure 1. Schematic of two general classes of problems.

Figure 1a corresponds to closed current loops comprised of either true currents or displacement currents. Induced scalar fields do not exist so the gauge concept cannot apply.

Figure 1b corresponds to the case where induced scalar fields exist and gauge choices are required. The two gauge choices give A_S' and A_C', as illustrated.